2013/7

Robustness, outliers and Mavericks in network regulation

Per J. Agrell and Pooria Niknazar

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium
http://www.uclouvain.be/core

CORE DISCUSSION PAPER
2013/7

Robustness, outliers and Mavericks in network regulation

Per J. AGRELL [1] and Pooria NIKNAZAR[2]

January 2013

Abstract

Benchmarking methods, primarily non-parametric techniques such as Data Envelopment Analysis, have become well-established and informative tools for economic regulation, in particular in energy infrastructure regulation. The axiomatic features of the non-parametric methods correspond closely to the procedural and economic criteria for good practice network regulation. However, critique has been voiced against the robustness of best-practice regulation in presence of uncertainty regarding model specification, data definition and collection. This paper investigates the foundation of the critique both conceptually and by describing the actual state-of-the-art used in energy network regulation using frontier analysis models in Sweden (2000-2003) and in Germany (2007-). A principal component of the applied frontier regulation is the systematic use of outlier detection models to define homogeneous reference sets and to exclude maverick reports. We review two families of outlier detection methods in terms of their function and application using a data set from Swedish electricity distribution, illustrating the different types of outliers. Finally, the paper concludes on the role of outlier detection as a mean to implement regulation with higher robustness.

Keywords: regulation, energy networks, outlier detection.

[1] Université catholique de Louvain, CORE and Louvain School of Management, B-1348 Louvain-la-Neuve, Belgium. E-mail: per.agrell@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.
[2] Université catholique de Louvain, Louvain School of Management, B-1348 Louvain-la-Neuve, Belgium. E-mail: pooria.niknazar@uclouvain.be.

The authors thank Andrew Johnson for generous access to code. This text presents research results of the ARC project on Shared Resources in Supply Chains, initiated by the Wallon Region. The scientific responsibility is assumed by the authors.

1. Introduction

One of the more prominent applications of state-of-the-art benchmarking is in the regulation of natural monopolies in general and electricity and gas networks, in particular. Benchmarking studies applied to inform such regulation has considerable economic impact on firms and consumers alike. The theoretical and intuitive appeal of using best-practice rather than average-practice cost norms in the regulation is undisputed. Still, economic regulation affecting private and public firms deploying large investments for essential infrastructure provision to the society must fulfil the highest criteria with respect to feasibility and regulatory robustness. In this paper, we will review some of the critique voiced against frontier-based regulation. In particular, we will relate the conjectures of various sources to the actual practice of energy network regulation with respect to the systematic use of outlier detection techniques. Specifically, we aim at addressing three research questions: (i) what are the specific requirements for structural and behavioral robustness in regulatory applications? (ii) what are the effects of using multi-stage outlier detection, theoretically and in real data sets? (iii) what is the final impact on regulatory robustness of the application of outlier detection methods?

The paper makes three contributions to the literature. First, it provides a conceptual view on the importance, specific requirements and classification of outlier detection and treatment for DEA applications to regulation. Given the prevalence of such applications in practice and their practical and economic importance, the paper fills an important gap in the current literature on frontier regulation. Second, although there is some scattered work on suggested applications of DEA and SFA to energy network regulation, there are no scientific papers documenting how the regulators actually assure robustness in DEA modelling, calculation and interpretation. This paper provides thus empirical evidence that can be used as factual reference for researchers working with methodological development. Third, the paper as such provides a response to some of the published critique raised against frontier analysis applications in regulation. As such, it contributes to the scientific discourse in general on the role and limitations of Data Envelopment Analysis in the public sector.

The outline of the paper is as follows: In section 2 we offer a review of frontier analysis methods and regulatory regimes, followed by the model notation for DEA. In section 3 we discuss the criteria and challenges of using frontier analysis in regulatory applications. Section 4 provides a classification of the outlier detection approaches in statistics and regulatory economics, as well as a review of existing approaches for outlier detection in DEA. The actual practice in frontier analysis is documented through the short case studies of the electricity network regulation in Sweden (2000-2003) in section 5 and in Germany (from 2007) in section 6. Section

7 is devoted to a comparative analysis of the outlier detection methods in section 4 applied to real data for the Swedish case. The paper is closed with conclusions and a final discussion in section 8.

2. Literature review

2.1. Frontier analysis methods

Best practice or *frontier analysis methods* model the frontier of the technology and identify a subset of the reference set units to form the peers, the performance of which is to be emulated by the others. The use of frontier models in regulation has practical as well as methodological advantages. In practice, the absence of *a priori* assumptions on the functional form and the foundation on a limited subset of identifiable best practice peers, make the frontier methods well adjusted to judicially implementable incentive regulation. Moreover, the behavioral effects defining attainable, yet evolving and demanding, performance targets are also well established in practice.

The *frontier* is defined as the edge of the empirical production possibility set. In frontier analysis, each firm is being seen as a decision Making Unit (DMU) which uses some inputs to produce some outputs, services or goods. The projection of the individual firm's position onto the efficient frontier determines the scope and areas for necessary performance improvements in order to achieve best practice [**?**]. The frontier analysis informs both *static* and *dynamic* efficiency assessments, i.e. the incumbent efficiency differences for a given year and the productivity improvements over time relative to technological progress. Generally, there are two main approaches for estimating and modelling the frontier; parametric and non-parametric, as well as two fundamental paradigms related to the data generation process, i.e. deterministic and stochastic models.

Parametric models are defined *a priori* except for a finite set of unknown parameters, estimated from the data. Parametric stochastic models consider the possible random noise and efficiency distributions in the data. Stochastic Frontier Analysis (SFA) is a family of methods in this category developed by **?**] and **?**]. Deterministic parametric models do not consider the noise in the data and any variation in data is considered be information about the shape of the production possibility set and, by extension, about firm efficiency. Corrected Ordinary Least Squares (COLS)[**?**], estimating a deterministic frontier using OLS is the main method of this type.

Non-parametric models relax the assumption of a known functional form. Of more limited attention in the regulation literature concerning stochastic nonparametric models we find the Stochastic Data Envelopment Analysis (SDEA) [**?**]. SDEA essentially relaxes the strict inclusion of all observations in the empirical production set in favour of a 'fuzzy' stochastic frontier [**?**]. A recent addition to the family of estimation techniques used for energy regulation came in 2011 for Finland, where the regulator proposed to use the StoNED method [? ?]. For the deterministic case, finally, Data Envelopment Analysis (DEA) constructs a piece-wise linear hull (envelope) around the empirical production set, based on linear programming. In the following section, we explain the DEA model in detail.

2.2. DEA

Expanding early work in ?], the name Data Envelopment Analysis (DEA) and the popularity of the approach were launched with the classical work in ?] and ?]. Below we make a condensed overview over relevant models and notation for our presentation, for a general description of the various DEA models, see texts such as ?] or ?].

The bearing principle of DEA is to construct a piecewise linear approximation of the best practice production set T from the observations using linear programming without requiring any imposed functional relationship between inputs and outputs. Following the convention, the observations are denoted Decision Making Units (DMU). DEA estimates the technology set T from the observed data on actual production activities based on the *minimal extrapolation principle*. The efficiency measure used in conventional DEA is a radial projection from the DMU to the efficient (best practice) frontier, either over inputs or outputs. Accordingly, the *efficiency frontier* is composed of those DMU classified as fully efficient.

To formalize the above, we assume that each of n DMUs, say DMU^i transform m_x controllable inputs x^i and m_z non-controllable categorical inputs z^i into m_y outputs y^i. The prices, if existing, on the controllable inputs are $w^i \in \mathbb{R}_+^{m_x}$.

We assume that the technological possibilities are the same for all DMUs' (except for the differences captured by the non-controllable variables). Specifically, these possibilities may be thought of as the set T of feasible input -output combinations

$$T = \{(x, z, y)|(x, z) \text{ can produce } y\} \tag{1}$$

We shall generally assume that T satisfy

Condition 1. *Free disposability:* $(x, z, y) \in T, x' \geq x, z' \geq z, 0 \leq y' \leq y \Longrightarrow (x', z', y') \in T$.

Condition 2. *Convexity: T is convex.*

Condition 3. *r returns to scale,* $(x, z, y) \in T \Longrightarrow (qx, z, qy) \in T, \forall q \in K\,(r)$ *,where* $k = "crs", "drs," \text{ or } "vrs"$, *and* $K(crs) = \Re_0, K\,(drs) = [0, 1]$ *and* $K\,(vrs) = \{1\}$, *respectively.*

For completeness, let us mention the non-convex model Free Disposability Hull (FDH) by ?], in which the only the first condition applies. As highlighted in

?] it is important to examine whether the underlying technology exhibits non-increasing, constant, or non-decreasing returns to scale. We leave the testing problem for scale returns in this context.

Given n observations of feasible production plans (x^i, z^i, y^i) the technical input-efficiency for a DMU facing non-controllable inputs z is $E^{TE}(x, y, z) : \mathbb{R}_0^{m_y} \times \mathbb{R}_0^{m_z} \times \mathbb{R}_0^{m_x} \to \mathbb{R}_0$ defined as

$$
\begin{array}{rll}
E^{TE}(x, y, z) = & \min & \theta \\
 & \theta, \lambda & \\
 & s.t. & \theta x \geq \sum_{i=1}^{n} \lambda^i x^i \\
 & & z\lambda^i \geq z^i \lambda^i \\
 & & y \leq \sum_{i=1}^{n} \lambda^i y^i \\
 & & \lambda \in \Gamma(r)
\end{array}
\tag{2}
$$

where $\Gamma(crs) = \Re_0^n, \Gamma(drs) = \left\{\lambda \in \Re_0^n | \sum_i \lambda^i \leq 1\right\}, \Gamma(vrs) = \left\{\lambda \in \Re_0^n | \sum_i \lambda^i = 1\right\}$. The second constraint effectively sorts the observations using the categoric variable z, [**?**]. The radial technical input-efficiency E^{TE} can be interpreted as the lower bound for the proportion of necessary input θx to achieve the observed output y under conditions z. A DMU is technically input-efficient if and only if the corresponding score $E^{TE} = 1$.

The associated underlying cost model for a DMU is given by

$$
C(y|z, w) = \min_x \{wx | (x, z, y) \in T\} \tag{3}
$$

The *DEA based cost norm* for a DMU facing input costs w and non-controllable inputs z is then$C^{DEA}(.|.,.) : \mathbb{R}_0^{m_y} \times \mathbb{R}_0^{m_z} \times \mathbb{R}_0^{m_x} \to \mathbb{R}_0$ defined as

$$
\begin{array}{rll}
C^{DEA}(y|z, w) = & \min wx & \\
 & x, \lambda & \\
 & s.t. & x \geq \sum_{i=1}^{n} \lambda^i x^i \\
 & & z\lambda^i \geq z^i \lambda^i \\
 & & y \leq \sum_{i=1}^{n} \lambda^i y^i \\
 & & \lambda \in \Gamma(r)
\end{array}
\tag{4}
$$

where $\Gamma(crs) = \Re_0^n, \Gamma(drs) = \left\{\lambda \in \Re_0^n | \sum_i \lambda^i \leq 1\right\}, \Gamma(vrs) = \left\{\lambda \in \Re_0^n | \sum_i \lambda^i = 1\right\}$. The DEA based cost function gives the minimal cost of producing the output for any output vector given the local factor prices and the local non-controllable conditions. A radial score E^{CE} is easily obtained through

$$
E^{CE}(x, y, z|w) = C^{DEA}(y|z, w)/wx \tag{5}
$$

To develop the setting into a full regulatory model we shall make some behavioral assumptions as well. Assume that the DMU's actual cost in the planning period is the minimal cost $C(y|z, w)$ plus whatever slack $s \in \mathbb{R}_0$ is introduced in the production process, i.e.

$$c(y) = C(y|z, w) + s \tag{6}$$

Note that production slack is summarized here as an additional cost, i.e., it is one-dimensional. The DMU (agent) knows $C(y|z, w)$ but the regulator (principal) does not. She does, however, know the input and outputs in n feasible (historical or inferred) production plans. Drawing on the information from the n production plans (x^i, z^i, y^i) the regulator can infer from the minimal extrapolation property [**?**] of the DEA model that

$$C(y|z, w) \leq C^{DEA}(y|z, w) \quad \forall y, z, w \tag{7}$$

The regulator has no more certain information about the cost structure than that expressed by the cost data collected, we are in a situation of *asymmetric information* concerning operating costs and conditions. The objective of the regulator is to induce the DMU to produce an exogenously given demand y at minimal cost under voluntary participation, as in e.g. **?**].

3. Frontier-based regulation

A number of different regulatory regimes have been applied to regulate natural monopolies; Rate-of-return regulation [**?**], cost-plus regulation [**?**], CPI-X revenue caps [**?**] and yardstick competition [**?**]. The predominant regime during the first phase of deregulation was to the high-powered regulation such as CPI-X, e.g. in countries such as England [**?**]. **?**] summarize some shortcomings of the CPI-X model such as the risk of bankruptcy, low cap and the risk of excessive informational rents for a loose cap. The yardstick regime gives some modifications to address the shortcomings of the CPI-X model while the inability to accommodate changes in outputs along with lack of dynamics still holds.

In order to guarantee long-term sustainability, a regulation regime should safeguard the owners' interests in investing in the service expansion and improvement, while simultaneously promoting efficient operation of the sunk assets. The objective is particularly clear in the regulation of energy network provision, such as electricity or gas transmission and distribution system operations. The terms *incentive regulation* (EU) or *performance based regulation (PBR)* are used to describe an approach where the any cost savings (or overshoot) are partially shared with the tariff payers. Recently, European countries have used the incentive regulation combined with advanced cost function estimation for the regulation of electricity and gas networks, both in distribution and transmission. The regulatory cost norms may be set relative to average or best practice frontier. Due to behavioral

Table 1: Some European regulation regimes and cost function methodologies for electricity distributors (DSO) and transmission operators (TSO). Participation in benchmarking at a nat[ional] or int[ernational] level without direct implementation in regulation is denoted by *.

Country	Regime	Method DSO	Method TSO
Austria	Revenue cap	DEA(nat)	DEA(int)*
Belgium	Revenue cap	DEA(nat)	DEA(int)
Denmark	Revenue cap	COLS(nat)	DEA(int)
Estonia	Revenue cap	COLS(nat)	DEA(int)*
Finland	Revenue cap	StonED(nat)	DEA(int)
France	Cost recovery	Ad hoc	DEA(int)*
Germany	Revenue cap	DEA-SFA(nat) best-of	DEA(int)
Great Britain	Revenue cap	COLS(nat)	DEA(int)*
Greece	Cost recovery	Ad hoc	DEA(int)*
Hungary	Price cap	Ad hoc	Ad hoc
Iceland	Revenue cap	Neg DEA(int)*	DEA(int)
Ireland	Price cap	Ad hoc	Ad hoc
Italy	Revenue cap(opex)	common X	DEA(int)*
Lithuania	Cost recovery	Ad hoc	DEA(int)*
Luxemburg	Cost recovery	Ad hoc	DEA(int)*
Netherlands	Yardstick	COLS(nat)	DEA(int)
Norway	Yardstick comp	DEA(nat)	DEA(int)
Portugal	Revenue cap	SFA(nat)	DEA(int)
Spain	Revenue cap	Engineering	DEA(int)*
Slovenia	Price cap	DEA(nat)	Ad hoc
Sweden	Rate-of-return	Ad hoc DEA(nat)*	DEA(int)*
Switzerland	Cost recovery	Ad hoc DEA(nat)*	Ad hoc

and economical advantages, the latter approach is dominating and the primary subject of our attention in this paper.

The international review by **?**] shows that some general advantages of DEA and SFA have led to the introduction of these benchmarking approaches in utility regulation. Table 1 below gives a summary of the benchmarking methodologies used for electricity DSOs in 15 European countries [**?**] with our updates for the period after 2008. Dynamically, the progression seems to be from more heavy-handed cost recovery regimes, passing through a period of model-based price fixation towards a high-powered market-based yardstick regime.

We see how some countries, like Spain and previously Sweden until 2006, have chosen to rely on technical engineering norms, sometimes referred to as ideal networks, in an attempt to identify not only *relative* best practice, but *absolute* technological possibilities. Most countries rely on some revenue cap model and have derived general productivity and individual inefficiencies using benchmarking tools like DEA and SFA. It is important to note that DEA, in particular for the transmission operations, has a widespread application throughout Europe, primarily through studies such as e^3GRID [?]. However, as indicated in Table 1 with a star in the column, many of the countries use DEA to inform regulatory rulings, but without direct link to the regulation. Nevertheless, the application of DEA is not without controversy. Therefore, in the following section, we review some of the the critique raised against the use of frontier-based models as DEA in regulation.

3.1. Robustness

The frontier analysis methods provide sounds empirical estimates for the cost function $C(y|z,w)$ as to inform regulatory proceedings, including tariff reviews, monitoring of cost and investment development as well as productivity development in the sector. The type of estimates most usually used are cost efficiency metrics, based on either best-practice (frontier) or quantile/average practice estimates. Since economic regulation is a judicial process that be can challenged in court if subject to bias or undue process, what matters in practice is not only the *estimation accuracy in expectation* but, at least as much, the *estimation process.* The objectives for frontier estimates concern both the efficiency model and the estimation robustness. For the efficiency model to be acceptable, the scope must be relevant to the data generation process, in particular the absence or presence of stochastic data as inputs or outputs in relation to the type of model chosen[1]. Second, the model scope must adequately reflect *multi-output services* unless they are perfectly correlated[2]. Third, the model specification must assure *structural comparability* as to enable the determination of a homogeneous reference set. In practice, this requirement means that the data collection and variable specification should be defined such that it is neutral to different accounting reporting standards, financial policies and non-regulated businesses.

The robustness of the model specification can be validated by crossvalidation using alternative specifications and even different methods. In the case the results are unique to a specific specification, the dependency should be carefully documented and evidence for the cost causality should be determined.

The robustness of the process implies that identical estimates should be ob-

[1]E.g., capacity investment and maintenance costs may be correlated to climatic conditions. Using annual realizations of climatic events in a model is equivalent to transforming it to a stochastic model. However, the causality with investments is not driven by annual realizations, but the expected climatic conditions, which can be used in a deterministic model.

[2]E.g., distribution of electricity at high- and low-voltage levels are two different services. If only one variable is used, operators with relatively higher/lower incidence of high-voltage to low-voltage delivery will be penalized/rewarded.

tainable with different software, analysts and advisors involved. The assumptions involved in the estimation should be a minimal set of cautiously determined parameters. The avoidance of technical parameters increases procedural reliability, i.e. variability due to interpersonal differences in skills, methodological assessment and non-verifiable assertions. The cautiousness in the parametrization is intrinsic in economic regulation since it is imposed on sunk investments that could be subject to opportunistic hold-up.

Finally, the robustness of the estimation itself relies on the assurance of a relevant reference set consisting of reliable data for structurally comparable operators. The next section will develop this aspect further by revisiting the critique against DEA in regulation.

3.2. Critique

The theoretical results by [**? ? ? ? ?**] referred to above show that DEA is optimal in regulation under mild conditions, including asymmetric information and multi-period regulation. The relatively widespread adoption of the method in applied utility regulation (cf. **?**] for an application to the German network regulation) provide evidence about the adequacy of the method to the real conditions in network regulation. Notwithstanding, as any regulatory method based on restricted discretion (or model-based regulation), the application cannot be made mechanically and care must be taken at several important steps in the model development, execution and interpretation.

The main point of criticism against use of the frontier-based models in regulation focuses either to *model specification errors* or *data uncertainty.*

Parametric models in general are robust to data errors, but sensitive to errors in the model specification. On the other hand, non-parametric models are sensitive to data errors, since they are based on a deterministic framework, but relatively robust to errors in model specification. The latter strength in terms of absence of non-verifiable *a priori* functional assumptions, partly explain the popularity of the non-parametric techniques in regulation where the endogenous frontier shape is a judicial strength.

Data errors are common, in particular in new and poorly monitored applications. **?**] state that poor quality data including incomparable data due to differing definitions of variables and missing data are the main source of problem with benchmarking models. **?**] claims that despite the claimed advantage by **?**], the use of DEA in regulation is a source of dispute and uncertainty. **?**] argues that any assumption by DEA such as "Productive inefficiency offers potential for cost reduction can" can be practically infeasible or any claims by DEA is in practice of regulation no more than assertions unsupported by analysis or evidence. Further, **?**] adds that any DEA model specification is subjective and may be contradicted by other, equally plausible models. The *choice of variables and model assumptions* (orientation, return to scale and convexity) and *selection of reference sets* are two obvious areas where the regulator can enter subjective preferences into the DEA model. A distortion in model specification may be unintentional, due to lack of skills, or intentional, in order to bias the results.

Another criticism against frontier-based models is the uncertainty which can result from unintentional data errors, inadequate variable specification on data collection techniques and intentional strategic misreporting of data (maverick reporting). In the presence of uncertainty in the data, the frontier's shape in DEA could change dramatically since DEA models, by construction, are sensitive to extreme values and outliers [**?**]. This heterogeneity in the reference set can also cause *lack of frontier robustness* which critics use to put the frontier- based regulation into question. The lack of robustness from data sensitivity can be seen in two perspectives: *structural* and *behavioral.*

Lack of robustness from a *structural perspective* usually comes from three main causes, cf. **?**]. First, the use of heterogeneous and incompatible technologies in the same reference set, second, operation at widely different scale in the data set and third, operation at different scope of activity, part of which may not be included in the model. All these reasons can make the data gathered from one firm incomparable to that of others if it uses a different technology, scale or scope. Models developed based on benchmarking with frontier models such as DEA which do not consider these differences can potentially put the financial stability and operating viability of a utility in jeopardy. **?**] notes that without a long history of standardization, data on regulated networks can be compiled and presented in different ways that make the comparisons incorrect. Indeed, a major part of the empirical work in international benchmarking such as **?**] is devoted to data harmonization and validation.

Lack of robustness from a *behavioral perspective* can be caused by either *collusion* or *maverick reporting* .

One of the main criticisms against benchmarking from a game-theoretic angle is that firms may have strong incentives to manipulate the regulatory process through *collusion.* The regulator using a frontier based model is dependent on information supplied by the firms. Collusion allows firms to behave as merged entities manipulating the frontier towards which all firms are measured. **?**] states that when regulated firms realize that they are played out against each other, they take precautionary measures to collude. Under collusive behavior aggregated data is correct but disaggregated data is infeasible. Theoretically, this would preclude any regulation of e.g. the yardstick type. However, the neutralization of the data collection would require an effective implementation of the collusive agreements. Collusions are implemented through monetary side payments, repeated contracting or reputation means. Normally, there is very limited financial interaction among

Table 2: Sources and intention levels by type of error in frontier-based models in regulation

Source of uncertainty	Strategic intention level Low	High
Firm	Reporting errors	Maverick reporting
Regulator	Inadequate variable specification	Biased model design

regulated firms, each operating a distinct network for a set of geographically located clients. Given that the regulator may audit and challenge any contract among regulated entities, side payments are difficult to arrange. Thus, none of the implementation instruments are readily applicable for a regulated utility.

Maverick reporting is the intentional misreporting in order to put the regulation model robustness into the question. In practice, when a new regulation regime is put in place, it cannot be predicted precisely how it will work. Based on generally acceptable standards for the quality of policy advice, it is assumed that the regulation regime is robust [**?**], but the regime might not work in the intended manner under some special circumstances. It is in the interest of some firms to provoke a "stress test" for the newly adopted regulatory model. The idea behind maverick reporting is to purposely produce absurd data in order to create unrealistic peers and to perturb the regulatory process. After the regulatory authority produces its ruling, the maverick firm will retrieve its data claiming mistakes in the data production procedure. Being based on frontier observation, the model estimation may have to be completely redone, provoking an instance of regulatory instability and high direct and indirect regulatory costs. From a legal perspective, unintentional misreporting is not considered fraud, in particular when the reporting firm manifests its "good will" through a submitted correction of data. Consequently, since the intention behind the reporting in unverifiable in court, regulatory provisions and penalties for this type of misreporting will not be enforceable in court. Hence, the possibility of maverick reporting will reduce the credibility and robustness of the model into question. The more centrally the maverick is placed, the more operators are affected by it as an incorrect peer.

Table 2 summarizes four possible causes of lack of robustness with respect to the origin of data uncertainty and the intention of the reporter.

For the regulator, the objective is to minimize the risk of regulatory failure which is associated with high social and economic costs [**?**]. Therefore, within an experimental framework, the regulatory authority tries to improve the best practice regulatory regime through a process of continuous improvement. Based on the criticism explained above, the focus is on those attributes that contribute to the robustness of the proposed frontier models and to the durability of the process regulatory regime.

Whereas regulatory benchmarking provides an effective and systematic basis of evidence of feasible production under various conditions, the discussion above shows that the stability of the model and its outcomes must prime on the precision of the model when using perfect data. This conclusion is a direct consequence of the regulatory application of frontier analysis and the observation that strategic action is always an option for firms under any type of regulation.

A principal instrument of frontier regulation to address this dilemma is the systematic use of outlier detection models to define homogeneous reference sets and to exclude inconsequential data. In the next section, we review the actual outlier detection methods used in network regulation in terms of their function and application in the regulation process, then the criticism against using frontier models in regulation are addressed. Ultimately, we will explain the role of systematic outlier detection in structural and behavioral robustness of frontier based models.

4. Outlier detection

Generally, there is no single definition of an *outlier* in the literature. We differentiate here between definitions that focus at the profile of the unit itself, *intrinsic definitions*, versus definitions that are based on the influence the unit exerts in a given estimation, *extrinsic definitions*.

According to a typical intrinsic definition, cf. **?**], an outlier is an observation which appears to be *inconsistent* with the rest of the data set. Inconsistent can be interpreted in different ways. An common statistical perspective denotes as outliers observations with *extreme values*. In the statistical sense, a point which lies three or four standard deviations from the mean is considered as an "outlier". One of the common reasons is that the observation could contain an error. Such outlier should ideally be corrected or perhaps be eliminated because it does not reflect a real production process. Another reason for an observation to be inconsistent with the rest of observations is the use of non-standard tools and procedures or being from a different data generating process (DGP). **?**] explain that if firms differ to a large extent from the rest of firms, they end up being badly captured by the model or having large impact on the model. The introduction of a new technology into a production process or non-observed environmental changes can also make a firm incomparable to others. In scientific or prospective analysis, the detection of profile outliers can be made *ad hoc* based on e.g. graphs, confidence intervals for predictions and technological data collected in a second stage. Classical profile criteria in frontier analysis are based on distance metrics $d(\theta_k(T))$ where the estimation depends on the production possibility set T. However, profiling is problematic in frontier-based regulation for several reasons. First, naturally the *ad hoc* identification of outliers is unacceptable in regulatory procedure. Given that the outlier is likely to be given a specific treatment, tariff or conditions, the procedure must be systematic, replicable and justified. Second, a more principal concern is that a regulatory authority cannot evoke profiling as a reason to exclude a firm from comparison. As opposed to a statistical estimation for research, where the analyst may be interested in some average expected relationships, regulation and taxation apply to all eligible subjects, irrespective of their profile. Excluding an operator is equivalent to revoking its eligibility, which must be justified by some other information, e.g. based on technological information or operating standards. A caveat to the same effect, but based on purely informational argument is found in the regression analysis literature, e.g. **?**]. Here, upfront rejection of data based on profiling arguments such as size, mix etc, is not advisable since many times these observations provide more, not less, information than the bulk of the sample.

According to an influence perspective, a unit acquires the quality of "outlier" for a given method and reference set through the (undue) impact that its inclusion gives on the quality of the estimation. Note the extrinsic quality of this definition compared to the previous intrinsic definition. In regression statistics an observation is considered inconsistent if it is substantially different from all other observations so that it induces large differences in the results of regression analysis [**?**]. With the statistical definition, an outlier denoted as "bad leverage" is a point situated far from the bulk of the points which the regression line crosses, so that it affect the slope of the regression line and reduce the precision of the regression coefficients. **?**] also define the *outlier* as an *influential observation*, which when is removed from the data set, parameter estimates change dramatically. In regression, the situation is somewhat more complex in the sense that some outlying points will have more influence on the regression than others. Standard regression diagnostic methods such as Cook's Distance [**?**], DFBetas, DFfits, and covariance ratios can be used to identify the possible influential outliers, see **?**] for a review.

In frontier models especially in *DEA*, *influential points* are considered outliers if they distort the frontier [**?**]. **?**] take the full step out and investigates "extreme efficient points" in DEA as outliers, without discussing their treatment. An outlier in a DEA model helps to span the frontier and may have a significant impact on the evaluation of several other firms. Although we can identify some isolated data points perturbed by noise as outliers, the outlier concept is different from the concept of noise in the data. Therefore, an analyst investigates the presence of both outliers and influential points because they can affect the DEA scores and change the whole regulatory procedure. We note here the difference between the extrinsic criteria when applied to parametric methods (regression) and non-parametric (DEA). In the former, the status of outlier is independent on the relative projection of the firm onto the frontier, opening for the identification of "inefficient outliers". In the latter, the mechanism of influence is uniquely passing through the determination of the frontier, which limits our attention to "efficient outliers" (cf. **?**]).

From an applied viewpoint, in particular in regulation, the restriction to efficient peers is attractive and useful for multiple reasons. First, the limited resources for data validation and auditing can be devoted to cases where potentially some useful public learning can occur, as opposed to the review of grossly inefficient firms, amounting to subsidized auditing services for underperformers.

Second, the incentive properties of a regulation in which a radical decrease in productive efficiency is associated with a positive chance of favorable treatment (i.e. classification as outlier) are poor. On the other hand, it is expected that claims for extreme productive efficiency would trigger a review to assure the reproducibility of the event.

4.0.1. Outlier detection methods in DEA

A number of approaches have been proposed in the literature to address outliers in DEA, among which the most widely cited are *super efficiency*, *log-ratio* and *order-m* methods. In the *super efficiency* **?**] and **?**], an observation which is significantly pushing out the frontier is considered an outlier. In the *log-ratio* method [**?**], an outlier is an atypical observation or a data point when removed from the data, the volume of a defined "data cloud" decreases more than if we remove any other observation. **?**] present a technique to classify for a further analysis those sample observations considerably affecting the measured efficiency score for the remaining units by checking whether these observations are contaminated by data errors or not. Pastor's technique allows determining when efficiency changes due to the presence of a given unit in the sample are statistically significant. In the *Order-m* method, **?**] propose a nonparametric estimation that can be applied in very general settings with multiple inputs and multiple outputs. The basic concept of Simar's method is the "expected frontier of order-m", where m is a "trimming" parameter of the frontier but also has its own empirical economic interpretation. In Simar's method, the estimator does not envelop all the observed data points so it is more robust to the extreme values. This method is used first as an exploratory data analysis, before using any frontier estimation. Simar's results are based on **?**] concept which proposes a non-parametric estimator, using the expected minimum input function of order-m. Cazals method is using the DEA nonparametric envelopment estimators by choosing m appropriately as a function of the sample size n, the estimator of the frontier, recovers the asymptotic properties of the FDH estimator. In the *order-m* method, an *outlier* is an atypical observation or a data point which for the different non-parametric estimator of m has the efficiency score of more than one which signifies that the observation is outlying the cloud of data set. Also, **?**] develop a unified model to identify both efficient and inefficient outliers in DEA. The measurement of outliers in Chen and Johnson's model is relative to a set constructed consistent with a subset of DEA axioms, ranking the individual outliers based on their influence on the measures.

Given the specific requirements in regulation, i.e. absence of technical parameters and outcomes, we focus our attention below to the method used in practice, super-efficiency filtering, as well as one example of the statistical methods, the log-ratio method by **?**].

5. Network regulation with DEA: Sweden

The Swedish electricity generation, transmission and distribution sector was unbundled and deregulated in 1996 following the first European energy market directive 96/92/EC. An independent regulator was launched, the Swedish Energy Authority (STEM)[3] with the mission to monitor the market and enforce cost efficient tariffs through an *ex post* regulation regime. To enforce the policy, STEM collected a wide range of data from the 245 concession area holders beginning from 1996. In 2000, two DEA models were developed to inform the regulator in the monitoring, initially used to provide public reports with easily color-coded scores (red, yellow, green) available to firms and customers (see e.g. **?**]). Below, we define these models, partially documented in **?**], relate to the observation of maverick reporting and summarize the initial provisions for DEA model robustness.

5.0.2. Model specification

The regulator desired to monitor both long-term efficiency using a technical efficiency model and short-term efficiency using a cost efficiency model. The *technical efficiency* (TE) model in Figure 1 shows the model specification. The outputs are total delivered energy (MWh) for low voltage and high voltage, respectively, the total number of connections for low and high voltage, respectively, and the total system peakload (MW). The inputs are network capital (SEK), total energy losses (MWh) and total network length (km) as a proxy for environmental conditions.

The *cost efficiency* (CE) model in Figure 2 corresponds to a typical cost function used in regulation, here with total expenditure (TOTEX, SEK) including depreciation, capital and operating costs, as the single input and the outputs retained from model TE above.

Some descriptive statistics are given in Table 3.

[3]Name changed in 2005 to *Energimarknadsinspektionen*, Swedish Energy Markets Inspectorate.

Table 3: Descriptive statistics on 126 Swedish electricity DSOs in 2000

Variable	unit		Min	Average	Max	Std.dev
Financial measures		MSEK				
	Revenues		2,701	64,367	1,354,918	132,940
	RAB[a] (book value)		0	212,316	7,985,024	725,396
Inputs				*		
	Total expenditure (TOTEX)	MSEK	2,684	54,275	993,529	100,104
	Energy losses	MWh	1	16,125	258,224	28,418
	Network capital	MSEK	0	212,316	7,985,024	725,396
	Total network length	km	113	1,293	8,842	1,340
Outputs						
	Total delivered energy LV	(MWh)	10	256,947	432,0286	46,2904
	Total delivered energy HV	(MWh)	1	13,0680	2,836,759	309,147
	Total number of connections LV	pc	768	20,724	451,398	46,200
	Total number of connections HV	pc	1	26	348	42
	Total system peakload	MW	2	2,278	77,380	10,920

[a]Regulatory asset base.

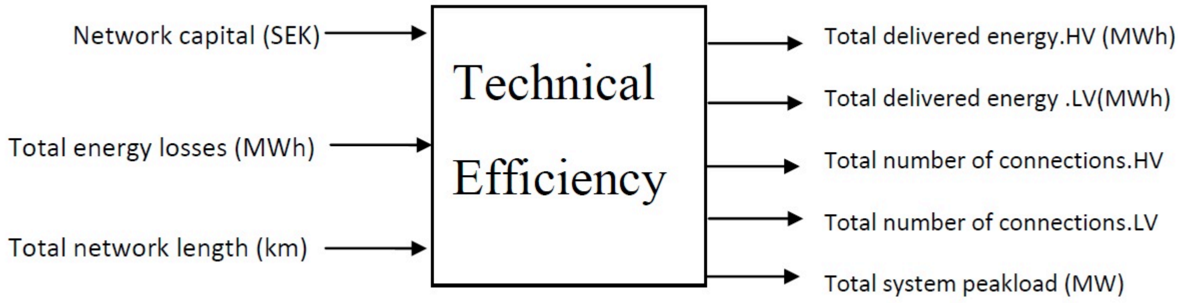

Figure 1: Technical input-efficiency (TE) model, electricity distribution [**?**].

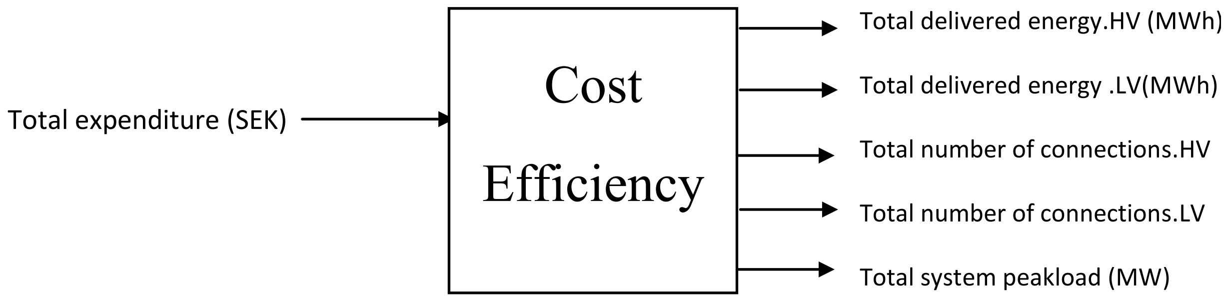

Figure 2: Cost input-efficiency (CE) model, electricity distribution [**?**].

5.0.3. Maverick reporting

Naturally, the electricity distributors in 2000 and 2001 had an important efficiency improvement potential, after years of low-powered regulation and passive owners. The results from the efficiency assessments in 2000 and 2001 in Table 4 confirm earlier results in e.g. **?**] that privately owned operators are more technically and cost efficient than municipally owned operators. However, it is interesting to note that the previously state-owned firm Vattenfall stand out. Not only is it the largest non-public utility in Sweden, facing a privatization and international expansion, it also appears as inefficient as the municipal firms.

Facing criticism in 2000, Vattenfall opposes the DEA method, the scores and their publication. The arguments raised by the firm were largely those of consultants as **?**], i.e. the claimed instability and arbitrariness of the scores and the model. Interestingly, Vattenfall changes the cost allocation among their 37 concession areas, resulting in shifts in controllable operating and total expenditure. In consequence, the individual scores of the Vattenfall concessions also change, as well as the identity of the Vattenfall operators on the efficient frontier. It is beyond

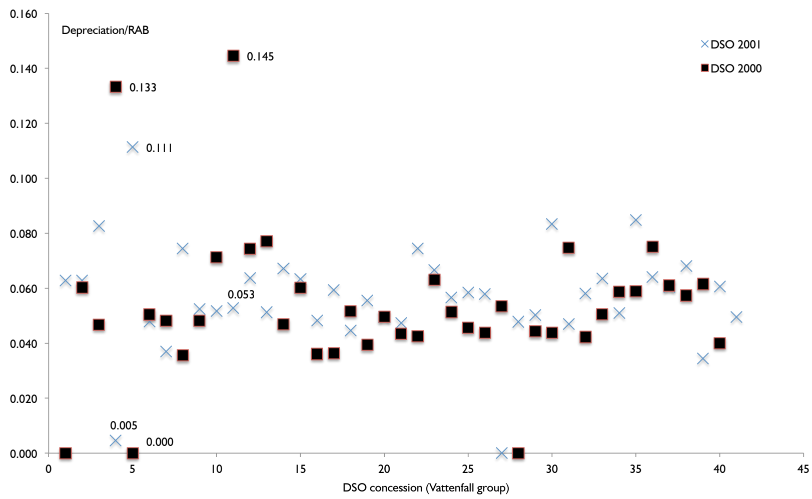

Figure 3: Depreciation ratio (depreciation/RAB) for Vattenfall DSO, Sweden, 2000 and 2001.

this paper to review in detail such changes, but the character of the action can be illustrated using a cost element that normally changes very little; the depreciation ratio for grid assets. In Figure 3 the concession areas operated by Vattenfall or its subsidiaries are listed alphabetic order and the depreciation ratios for the years 2000 and 2001 are depicted. Note that several units (e.g. areas 4, 5, 13) change policy drastically between the two years, with some units showing over 10 per cent ratio and some units less than 1 per cent. The aggregate depreciation ratio (5.47% in 2000, 5.68% in 2001) did not change for the Vattenfall group during the period. At their own request and the outlier review of the regulator, Vattenfall later corrected the operating cost data (including the allocation of depreciation). Given the context (an early introduction of DEA, incentives for the operator to discredit the method), the action could be interpreted as early maverick reporting.

5.1. Robustness implementation

Noting the early experiences with a large number of DMU, varying data quality and strategic reporting from operators controlling multiple concession areas, the regulator implemented a simple, yet effective DEA robustness model through the

Table 4: Cost efficiency scores by ownership, Swedish DSOs 2000, 2001. Balanced panel, 7 DSO excluded (only active in 2000).

Owner	Group	n^a	n	TOTEX[b]	TOTEX [c]	CE_{SR}	CE_{SR}	TE_{LR}	TE_{LR}
		2001	2000	2001	2000	2001	2000	2001	2000
Municipal	all	103	103	3,954,008	3,944,104	28%	28%	24%	27%
	corp.[d]	95	95	3,784,252	3,771,553	28%	28%	25%	29%
	utility[e]	8	8	169,756	172,551	35%	36%	2%	2%
Private	all	101	101	6,375,120	6,268,060	20%	18%	22%	26%
	Fortum	18	18	1,801,402	1,761,543	18%	11%	21%	23%
	Sydkraft	14	14	1,539,593	1,542,973	6%	4%	11%	18%
	Vattenfall	37	37	2,204,405	2,140,573	25%	26%	27%	33%
Cooperatives[f]	all	31	31	378,485	347,702	36%	42%	25%	28%
Total		235	235	10,707,613	10,559,867	24%	23%	23%	27%

[a] Balanced panel, 7 operators only operating in 2001 excluded, included cost in 2001 correspond to 98.9% of total cost.
[b] Total expenditure in kSEK excl taxes, energy for network losses and fees for transmission.
[c] Total expenditure in kSEK excl taxes, energy for network losses and fees for transmission.
[d] Incorporated firm, limited stock. Staff and management employed by independent structure.
[e] Public utility within the organization of a municipality, staff are municipal employees.
[f] Consumer cooperatives, non-profit associations, other

”peeling” technique in **?**][4].

The robustness of the DEA assessment is improved if the regulatory ruling is based on the ordinal classification rather than the exact radial projection, which is more sensitive to the peer unit data. Given that the initial regulatory model in Sweden was based on enforced *ex post* regulation, there was no need to jeopardize the model by providing too detailed assessments. Consequently, the regulator calculated the ranking in terms of ”peels” or groups using the following procedure:

1. Set the group number to $g = 0$
2. Let $\Omega(g)$ be the set of all n DMUs
3. Solve E^{TE} or E^{CE} using $\Omega(g)$ as reference set.
4. Increment $g := g + 1$
5. Define $\Psi(g) = \{i \in \Omega(g-1) : E(x^i, y^i, z^i) = 1\}$, i.e. the set of efficient DMU in group g
6. Reduce the reference set $\Omega(g) = \Omega(g-1) - \Psi(g)$
7. If $\Omega(g) \neq \emptyset$ then repeat from Step 3, else stop.

The resulting groups were translated for pedagogical reasons into colorcodes according to a rule based on the peers in groups 3 and 6, i.e. $\Psi(3)$ and $\Psi(6)$. A ”green” layer or group was defined as one where all DMUs are efficient with respect to the DMU in $\Psi(3)$. Likewise, a ”yellow” layer or group is such that all DMU are efficient with respect to $\Psi(6)$. Remaining layers are classified as ”red”. An example for 2002 is given in Table 5 below, taken from **?**].

The regulatory logic behind the coloring scheme is essentially an evidence-based argument for selective monitoring. Unless there are two independent sets of DSOs dominating a unit, the regulator did not deem the evidence strong enough to initiate a tariff review, fearing a backlash in case of appeal. However, all yellow units have between three and five independent sets of peers that dominate their performance, which must be considered a plausible robustness irrespective of potential errors or maverick reporting among the firms. Consequently, these firms were subject to selective reviews. Finally, for the ”red” firms for which more than six independent sets of peers could be formed to prove their productive inefficiency, there is an impelling argument for *ex post* tariff review for all operators in the class.

[4] The seminal work on this is reported as [**?**]

Table 5: Efficiency results for Swedish DSO ($n = 226$), DEA-CE, 2002. [?], Table 8.

g	Color	n_g	Average $E^{CE}(g = 0)$	n_s
1	Green	32	100%	2
2	Green	34	85%	13
3	Yellow	37	75%	6
4	Yellow	28	65%	5
5	Yellow	16	54%	5
6	Red	16	45%	7
7	Red	9	42%	2
	Grey	38		

Used without outlier detection, the classification would provide a false protection for firms operating at or reporting odd production profiles in order to achieve high DEA efficiency estimates. The regulator decided to label firms as "grey" if they were measured as fully efficient in a layer, but did not appear in the peer set for any firm in the reference set. In Table 5 there are 38 self-evaluators. The set of peers for any group g is partitioned in n_g "colored" units and $n_s = \|\Psi(g)\| - n_g$ self-evaluating or "grey" units.

6. Network regulation with DEA: Germany

The German regulation is basically a revenue cap regulation. Each regulatory period is 5 years and the content of the first two regulatory periods have been detailed, giving the DSO more long-term forecasts on which to act.

It is also a Totex based regulation, i.e., both operating expenses (Opex) and capital cost expenses (Capex) are subject to regulation. Capital costs are based on either book values or standardized costs using replacement values and constant annuity calculations of yearly cost using life times of different asset groups.

The revenue cap of an individual DSO k in the German regulation in year t is determined by the formula

$$R_t^k = C_{nc,t}^k + (C_{tnc}^k(0) + (1 - V(t))C_c^k(0))(\frac{RPI(t)}{RPI(0)} - x(t))ExFa(t) + Q(t) \quad (8)$$

where C_{nc} is the cost share that cannot be controlled on a lasting basis (statutory approval and compensation obligations, concession fees, operating taxes etc.), C_{tnc} is the cost share that cannot be controlled on a temporary basis (essentially the efficient cost level found as the total costs multiplied by the efficiency level, C_c are the controllable costs, $V(t)$ is a distribution factor for reducing inefficiencies (initially set to remove incumbent inefficiency after two regulatory periods, i.e., 10 years), RPI(t) is the retail price index in year t, $RPI(0)$ is the retail price index in year 0, and $x(t)$ is the general productivity development from year 0 to year t reflecting the cumulative change in the general sectoral productivity factor for year t of the particular regulatory period relative to the first year of the regulatory period. Also, ExFa is an expansion factor reflecting the increase in service provision in year t compared to year 0 and determined as

$$ExFa_j^k(t) = 1 + max(\frac{L_j^k(t) - L_j^k(0)}{L_j^k(0)}, 0) \tag{9}$$

where $L_j(t)$ is the volume of load at level j in year t of the particular regulatory period. The expansion factor for the entire network is the weighted average of all network levels. Lastly, Q(t) is the increase or decrease in the revenue cap from quality considerations. Revenue caps may have amounts added to or deducted from them if operators diverge from required system reliability or efficiency indicators (quality element). The quality element is left to the discretion of the regulator.

6.1. Robustness implementation

From a benchmarking perspective, the regulation is remarkable for being explicit with respect to a series of technical aspects such as cost drivers, estimation techniques, return to scale and outlier criteria.

The Ordinance is specific about a *minimal set of cost drivers*. Cost drivers such as connections, areas, circuit length, and peak flow, were obligatory. Of course, this leaves a series of available alternatives even within these groups and it does not exclude cost drivers covering other aspects of the service provision.

The German incentive regulation is also explicit as to which *estimation techniques* to use in benchmarking electricity and gas DSOs and how to combine the results of multiple models. According to Section 12 of the Ordinance, the efficiency level for a given DSO is determined as the maximum of four efficiency scores, $E_{DEA}(B), E_{DEA}(S), E_{SFA}(B)$, and $E_{SFA}(S)$, where E_{DEA} is the Farrell costefficiency, calculated with a NDRS-DEA model, E_{SFA} is the Farrell cost efficiency, calculated using a SFA model, and the argument B denotes book value and S standardized capital costs. As such, the regulation takes a cautious approach and biases the decision in favor of the DSOs in case of estimation risk. Entities demonstrating particularly low efficiency are given the minimum level of 60 percent. In summary, the efficiency of DSO k is calculated using the equation:

$$\max\{E_{DEA}^k(B), E_{DEA}^k(S), E_{SFA}^k(B), E_{SFA}^k(S), 0.6\} \tag{10}$$

The Ordinance is explicit outlier detection. Indeed, it prescribes two outlier criteria to be tested for each DSO, and if any of them is fulfilled, the DSO cannot be allowed to affect the efficiency of the other DSOs. The two criteria can be formalized in the following ways. Let Ω be the DSOs is the data set, and $k \in \Omega$ be a potential outlier. Also, let, $E(h, \Omega)$ be the efficiency of h when Ω is used to estimate the technology and let $E(h, \Omega \setminus k)$ be the efficiency when DSO k does not enter the estimation.

The *first outlier criterion* is that a single DSO should not have too large of an impact on the average efficiency. We can evaluate the impact on the average efficiency by considering

$$\frac{\sum_{h \in \Omega \setminus k}(E(h, \Omega \setminus k) - 1)^2}{\sum_{h \in \Omega \setminus k}(E(h, \Omega) - 1)^2} \tag{11}$$

The test compares the average efficiency of the other operators when k cannot affect the technology as compared to the average efficiency of the other DSOs when the k is allowed to impact the evaluations. Since $E(h, \Omega \setminus k) \geq E(h, \Omega)$, this ratio is always less than or equal to 1, and the smaller the ratio is, the larger the impact of k, i.e., small values of the ratio will be an indication that k is an outlier. The asymptotic distribution of the ratio is $F(|\Omega\| - 1, |\Omega\| - 1)$, see **?**].

The *second outlier criterion* is that no DSO k will be extremely super-efficient in the sense that

$$E(k, \Omega \setminus k) > q(0.75) + 1.5(q(0.75) - q(0.25)) \tag{12}$$

where $q(a)$ is the a quantile of the distribution of super-efficiencies, such that e.g., $q(0.75)$ is the super-efficiency value, below which exist 75% of DSOs.

In addition to these outlier rules, the ordinance prescribes the use of common econometric outlier detection methods like Cook's distance.

7. Outlier detection techniques at work

In this section, we review, operationalize and analyze the outlier detection techniques used or proposed for identifying outliers in frontier analysis methods in regulation. To concretely illustrate the properties and comparative performance for each of these methods, we use an illustrative example of Swedish electricity distribution system operators for the year 2000, previously discussed. The data set obtained from the regulator EMI includes a crosssectional dataset of technical and cost observations for 152 DMU, excluding 20 incomplete observations.

7.1. Super efficiency

What is now called super-efficiency and routinely calculated by most DEA software was first suggested by **?**][5] as a means to differentiate among frontier units. The formalization of superefficiency for outlier detection was made in **?**].

[5]The first mention of the principle for detection of outliers was made in **?**]

Let T^{*-k} be a DEA approximation of the technology based on all observations but DMU^k, i.e.

$$
\begin{array}{rl}
E^{SUPER}(k) = & \min \quad \theta \\
& \theta, \lambda \in \mathbb{R}_+^{n-1} \\
& s.t. \quad \theta x^k \geq \sum_{i \neq k} \lambda^i x^i \\
& \qquad z^k \lambda^i \geq z^i \lambda^i \\
& \qquad y^k \leq \sum_{i \neq k} \lambda^i y^i \\
& \qquad \lambda \in \Gamma(r)
\end{array}
\tag{13}
$$

Technically, it is simple to set up the associated mathematical programs — they look just like the usual one except that one column has been eliminated corresponding to the λ^k variable.

The super efficiency measures on the input and output sides are not restricted to be either below or above 1. Indeed, this is part of the motivation for them - wee need to be able to differentiate among the units with traditional efficiency scores of 1. The input super efficiency score $E^{SUPER}(k)$ may be larger than 1 with the interpretation that DMU^k could have increased its inputs with a factor $E^{SUPER}(k)$ and still not have been dominated by an feasible reference unit.

It follows also from the definition that the traditional efficiency measures are simply aggregates of the super-efficiency measures

$$
E(k) = \min\{E^{SUPER}(k), 1\} \tag{14}
$$

Hence, the super-efficiency measures contains at least the same and sometimes additional information. It is obvious therefore that they are advantage for decision making and incentive purposes - at least as long as we ignore information processing costs.

Definition 1. *In the superefficiency method, an outlier is defined as an observation k such that $E^{SUPER}(k) > \eta$ for a predefined threshold level $\eta > 0$.*

7.1.1. Applied example 1

The number of outliers in the superefficiency method is a function of the trimming factor η. Following the German regulation in section 6, we define η by so that no DSO k will be extremely super-efficient in the sense that

$$
E(k) = (q(0.75) + \eta(q(0.75) - q(0.25))) \tag{15}
$$

where $q(a)$ is the a quantile of the distribution of super-efficiencies, such that e.g., $q(0.75)$ is the super-efficiency value, below which exist 75% of DSOs. The calculation yields $\eta = 1.5$. Figure 4 also shows that around $\eta = 1.5$ the number

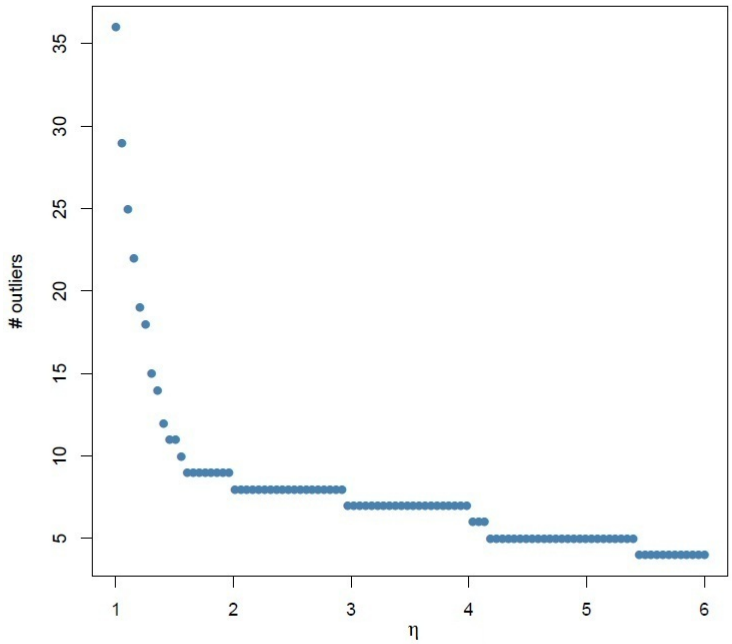

Figure 4: Number of outliers in the TE model for each trimming criteria

of outliers decreases from 35 outliers to an almost stable level around 10 outliers. Trimming factors of $1.5 < \eta < 3$ result in a relatively stable set of outliers.

In figure 5, we show the difference between the efficiency score for the non-outliers before and after outlier removal for the $TE - VRS$ model, fixing $\eta = 1.5$ and excluding the resulting 10 outliers from the reference set.

Naturally, Figure 5 confirms the increase in individual scores for the non-outliers at their removal. However, the average finite[6] super efficiency score drops from an unrealistic 467.4% to a more plausible score of 92.9% whichis around 1 and seems reasonable. .

7.2. Log-Ratio

The *Log-ratio* method, is a statistical methodology for identifying outliers in deterministic non-parametric frontier models by performing various sensitivity analyses or by deleting ostensibly efficient observations until efficiency estimates are stabilized [?]. This methodology is useful in identifying observations that may contain some form of measurement error and thus merit closer scrutiny. When data checking is costly, log-ratio methodology can rank the observations in terms

[6]Infinite superefficiency values are excluded from the calculation of averages.

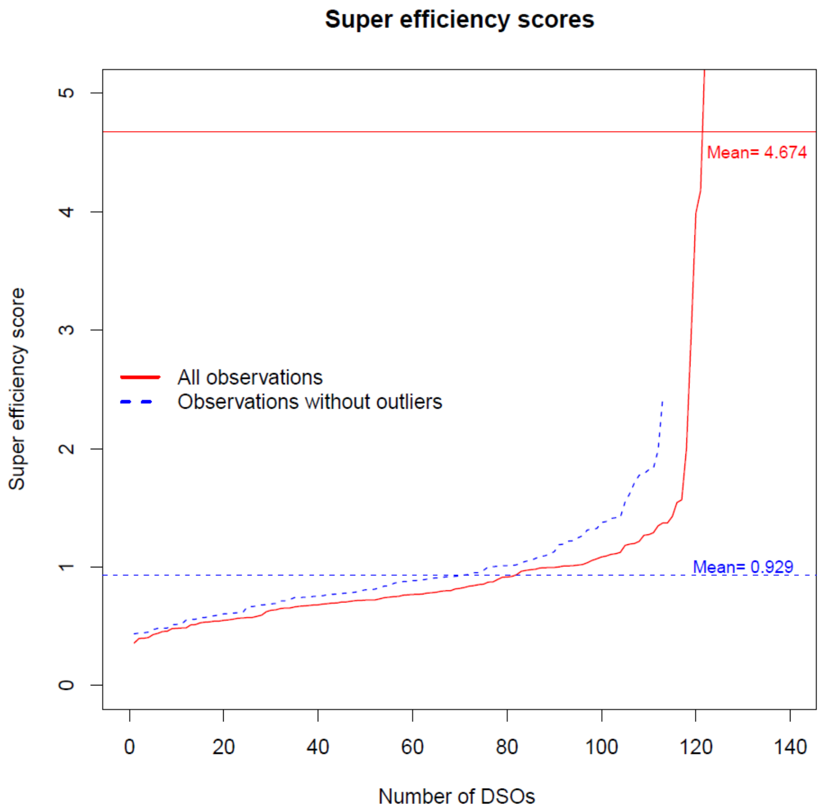

Figure 5: Difference between super efficiency scores with and without the outliers identified by the super efficiency method, $eta = 1.5$

of their dissimilarity to other observations in the data. It can define a priority for further inspection of the data set. In the log-ratio method, it is also possible to extend the definition of *outlier* in order to remove a group of observations and see the effects on the log-ratio measure in contrast to the super efficiency method that we should remove the observations one by one from the data set to see the effects.

The Log-Ratio method is based on a defined data cloud (D) and its volume [**?**].The main idea behind this method is that if we remove a firm from the data, then the volume of the data cloud decreases. The decrease in the data cloud can be used as an indication that the firm is an *outlier*. **?**] defines the *volume* of the data set as the determinant of the corresponding input-output matrix,$[x, y]$. Also, in order to look for a group of outliers, we can remove more observations and look at how much the volume of the cloud changes in consequence. One good example of using Log-ratio method in regulation was carried by **?**] by using this method for regulation of Brazilian electricity distribution utilities.

To identify the outliers, the Log-ratio method looks for small values of R^i which is the ratio of cloud of data without the removed firm on the cloud of data with all the observations. This method finds the smallest R for each number of firms that could be identified as outliers in the data set. Graphical tools in this method are used to identify the isolated points as an indication for the number of outliers that should be deleted from the data set.

Let $D^{(i)}$ be the determinant of matrix consisting the input and output data for all the firms after removing firm i. $R^{(i)}$ is the ratio between the new volume of the data cloud and the old volume of the data.

$$R^{(i)} = \frac{D^{(i)}}{D} \tag{16}$$

By evaluating the changes in $R^{(i)}$, we can also see the consequence of removing two or more firms from the data cloud. For example; we can use $D^{(\varpi)}$ as the *volume* for subset $[\Omega - \varpi]$, when $\varpi \subseteq \Omega$. To identify outliers or groups of outliers, this method looks for smallest values of R for each number of firms that are deleted from the data set. Along with that, a graphical method is used to plot:

$$\left(p, \log(\frac{R^{(p)}}{R^{(p)}_{\min}}) \right) \tag{17}$$

where p is the number of deleted firms. In the graph, the p with isolated low points gives an indication of p outliers which should be eliminated from the data set.

Definition 2. *In log-ratio method, an outlier is defined as a DMU k such that when it is removed from the data cloud (D), $R^{(k)} < \eta$ for a predefined threshold level $\eta > 0$.*

Table 6: The p removed observation corresponding to a minimum value of $R^{(p)}$ for TE model

p													$R^{(p)}_{\min}$
1	94												0.0296
2	32	94											0.0062
3	72	32	94										0.0029
4	125	105	32	94									0.0014
5	125	105	72	32	94								0.00067
6	110	114	88	44	32	94							0.00014
7	110	114	88	44	72	32	94						6,87E+09
8	125	110	105	114	88	44	32	94					3,32E+09
9	125	110	105	114	88	44	72	32	94				1,54E+09
10	125	110	105	114	88	29	44	72	32	94			8,65E+08
11	125	110	60	105	114	88	29	44	72	32	94		4,92E+08
12	125	110	60	105	114	88	42	29	44	72	32	94	2,75E+08

7.2.1. Applied example 2

With applying log-ratio method on our data set, the initial results identifying the outliers for the TE model are presented in Table 6. The rows in this table show which deletions give the minimum value of $R^{(p)}$; this minimum value is also given in the right-most column. Thus, the first row, $p = 1$, shows that deleting firm 94 from the data set results in a value of $R^{(p)}$ at 0.0296 and that this value is the minimum value of $R^{(1)}$ that means the minimum value of R when just one firm is deleted from the dataset.

Figure 6 shows the number of observations that should be identified as outlier. We can see that the dashed line peaks at 6 deleted firms($p = 6$). Therefore, we have 6 outliers. From Table 6, we can see that these firms are $94, 32, 44, 88, 114$ and 110.

In Figure 7, we show the difference between the efficiency score for the non-outliers before and after outlier removal for the $TE - VRS$ model. We have excluded 6 outliers from the data set as they were identified in Figure 6. The outliers are 5 peers and 1 almost efficient unit ($E^{TE} = 97.1\%$), on average 8 times larger in terms of number of connections (LV), but 110 times larger in terms of maximum peakload. Only two of the outliers figure in the peer sets of other units, 17 and 9 times, respectively. We can then conclude that the metric indeed manages to identify very specific units, but the data-cloud method does not capture the influence dimension.

Figure 7 confirms a marginal increase in the score for the non-outliers, the average TE-score goes from 77.47% to 78.39%.

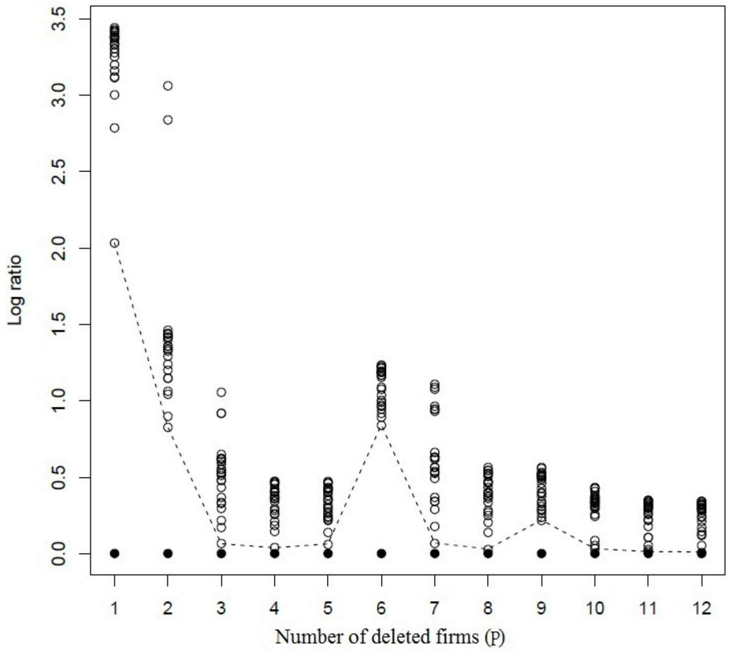

Figure 6: Number of outliers in mode TE for each trimming criteria

7.3. Comparison of the results of the outlier detection methods

In order to compare the outlier methods and to show how they affect the DEA results, we have considered the simple DEA model in both CRS and VRS technology and we compare the average DEA efficiency scores before and after removal of outliers by different methods. The comparison results are given in Table 7.

As we see in Table 7, the average efficiency scores increase if we remove the outliers. As noted, the differences in the average efficiency scores between the log-ratio method and the full sample are surprisingly low, in spite of removing six DMUs. However, the internal consistency is low, since there is no common

Table 7: Comparison between average DEA scores after removing the outliers by different methods

Technology	Average TE All observations	Super efficiency	Log-ratio
CRS	74.85%	77.07%	74.94%
VRS	78.49%	82.26%	78.72%
Number of outliers identified		10	6
Percentage of outliers		7.9%	4.7%

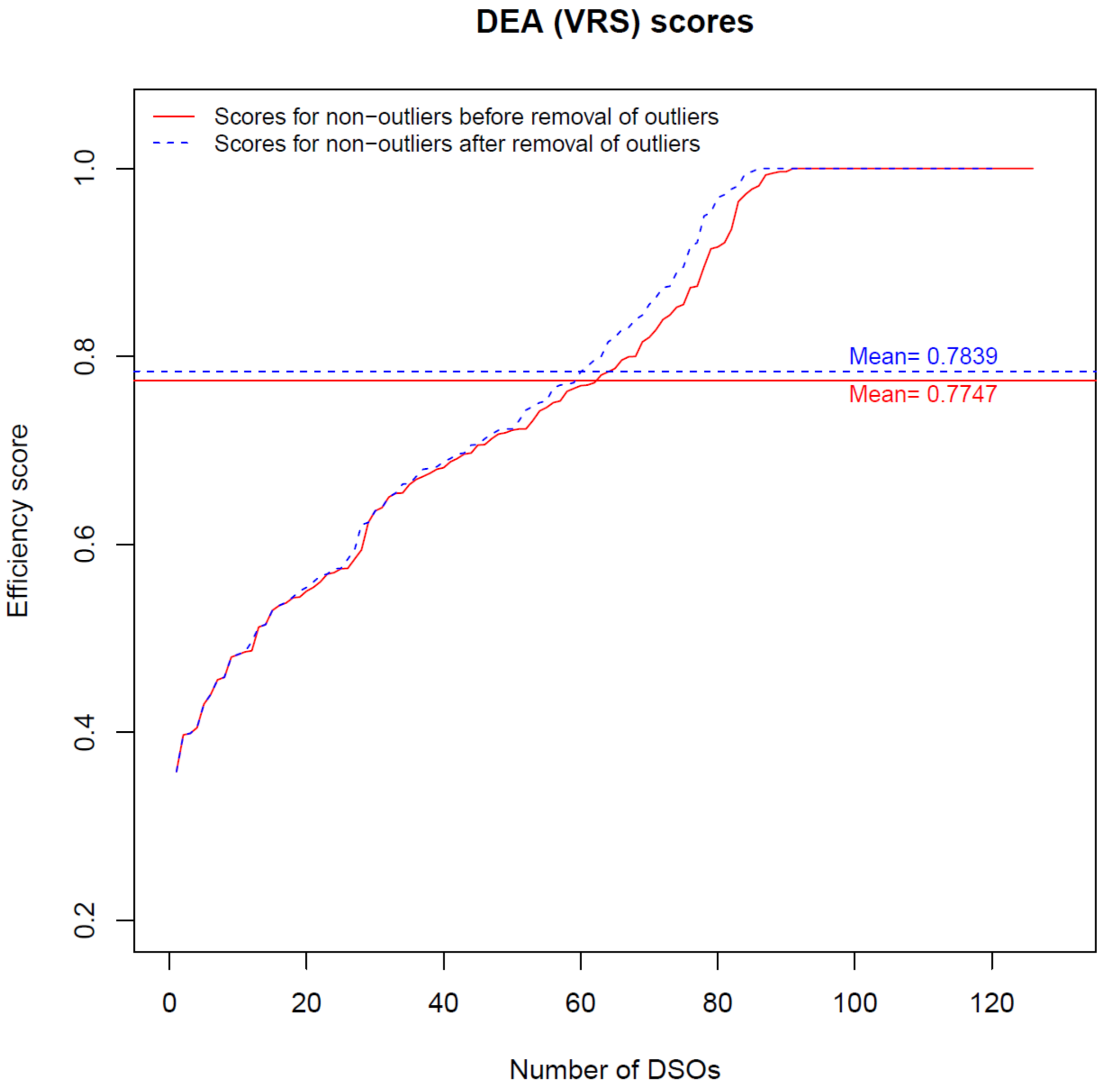

Figure 7: Difference in TE scores with and without the outliers identified by Log-ratio method.

observation identified as outlier in both super efficiency and the log-ratio method in this example.To relate back to the maverick anecdote in section 5, we note that

no Vattenfall operator was identified as outlier by the log-ratio method.

8. Conclusions

In this paper we have provided some evidence for the robustness of frontier-based methods for economic regulation. Lack of robustness may be caused by heterogeneity in the reference set, unintentional data errors, erroneous functional specification of the production frontier or intentional strategic misreporting of data (maverick reporting). Outlier detection is shown to be an effective filter for reference set heterogeneity and unintentional reporting errors, both occurring in actual examples of European network regulation. In the Critique section, four main drawbacks of employing frontier-based models in regulation process are explained. These points can now be addressed one by one through the paper:

First, employing systematic outlier detection is very useful in addressing data uncertainty which comes from unintentional data errors. In order to minimize problems due to data errors, a very careful analysis of data along with using multi stage outlier detection should be done prior to use the regulatory model on the data set and build the frontier. With identifying and removing exceptional and inconsistent data, the robustness of regulatory model will increase.

Second, in the regulatory process, firms most establish the appropriate reporting formats and standardization of data in order to ensure the regulator in the quality of data that have been reported. Therefore, the usage of non-standard tools and procedures can be detected as an outlier and therefore be separated from rest of the data set in the regulatory process. Multi-stage outlier detection is useful in achieving robustness in this kind of inconsistent data reporting.

Third, frontier-based models in regulation measure firms against each other and to the frontier so they can put a lot of restrictions on the side payments. A massive side transfer may be required from second agent two to the first agent in order to induce the first agent to truthfully reveal his type. Consequently, firms would prefer collusive agreement that does not depend on information exchange that make it hard for the firms to collude.

Forth, strategic misreporting of the maverick type challenges primarily mechanic and semi-automatic applications of frontier methods to determine cost norms. The threat would be most serious early in the life cycle of the regime, in situations of conflict concerning the model integration in regulation and for complex data collection. The review of current regulatory practice shows discretionary and 'soft' ranking approaches to be used in such situations, e.g. in transmission system benchmarking and in the introduction of DEA benchmarking in Sweden and elsewhere. By refraining from a precipitated use of the method in tariff regulation, the regulator as well as the firms may gain both confidence in and information from the model.

Recent titles

CORE Discussion Papers

2012/32 Jean-François MERTENS and Anna RUBINCHIK. Equilibria in an overlapping generations model with transfer policies and exogenous growth.
2012/33 Jean-François MERTENS and Anna RUBINCHIK. Pareto optimality of the golden rule equilibrium in an overlapping generations model with production and transfers.
2012/34 Hélène LATZER and Florian MAYNERIS. Income distribution and vertical comparative advantage. Theory and evidence.
2012/35 Pascal MOSSAY and Takatoshi TABUCHI. Preferential trade agreements harm third countries.
2012/36 Aitor CALO-BLANCO. The compensation problem with fresh starts.
2012/37 Jean-François CARPANTIER and Arnaud DUFAYS. Commodities volatility and the theory of storage.
2012/38 Jean-François CARPANTIER and Christelle SAPATA. Unfair inequalities in France: A regional comparison.
2012/39 Shui-Ki WAN, Shin-Huei WANG and Chi-Keung WOO. Total tourist arrival forecast: aggregation vs. disaggregation.
2012/40 Thierry BRECHET, Yann MENIERE and Pierre M. PICARD. The clean development mechanism in a global carbon market.
2012/41 Nikolaos GEORGANTZIS, Rafael MONER-COLONQUES, Vicente ORTS and José J. SEMPERE-MONERRIS. Theoretical and experimental insights on firms' internationalization decisions under uncertainty.
2012/42 Berardino CESI and Dimitri PAOLINI. Peer group and distance: when widening university participation is better.
2012/43 Arnaud DUFAYS. Infinite-state Markov-switching for dynamic volatility and correlation models.
2012/44 Miguel JARA, Dimitri PAOLINI and J.D. TENA. Management efficiency in football: An empirical analysis of two extreme cases.
2012/45 Guillaume ROELS, Philippe CHEVALIER and Ying WEI. United we stand? Coordinating capacity investment and allocation in joint ventures.
2012/46 Mikel BEDAYO, Ana MAULEON and Vincent VANNETELBOSCH. Bargaining and delay in trading networks.
2012/47 Pierre M. PICARD and Ridwan D. RUSLI. State owned firms: Private debt, cost revelation and welfare.
2012/48 Shin-Huei WANG, Luc BAUWENS and Cheng HSIAO. Forecasting long memory processes subject to structural breaks.
2012/49 Thierry BRECHET, Carmen CAMACHO and Vladimir M. VELIOV. Adaptive model-predictive climate policies in a multi-country setting.
2012/50 Vladyslav NORA and Hiroshi UNO. Saddle functions and robust sets of equilibria.
2012/51 Thomas BAUDIN, David DE LA CROIX and Paula GOBBI. DINKs, DEWKs & Co. Marriage, fertility and childlessness in the United States.
2012/52 David DE LA CROIX and Omar LICANDRO. The longevity of famous people from Hammurabi to Einstein.
2012/53 Marijn VERSCHELDE, Jean HINDRIKS, Glenn RAYP and Koen SCHOORS. School staff autonomy and educational performance: within school type evidence.
2012/54 Thierry BRECHET and Susana PERALTA. Markets for tradable emission permits with fiscal competition.
2012/55 Sudipto BHATTACHARYA, Claude D'ASPREMONT, Sergei GURIEV, Debapriya SEN and Yair TAUMAN. Cooperation in R&D: patenting, licensing and contracting.
2012/56 Guillaume WUNSCH, Michel MOUCHART and Federica RUSSO. Functions and mechanisms in structural-modelling explanations.
2012/57 Simone MORICONI, Pierre M. PICARD and Skerdilajda ZANAJ. Commodity taxation and regulatory competition.

Recent titles

CORE Discussion Papers - continued

2012/58 Yurii NESTEROV and Arkadi NEMIROVSKI. Finding the stationary states of Markov chains by iterative methods.
2012/59 Tanguy ISAAC and Paolo PIACQUADIO. Equity and efficiency in an overlapping generation model.
2012/60 Luc BAUWENS, Giuseppe STORTI and Francesco VIOLANTE. Dynamic conditional correlation models for realized covariance matrices.
2012/61 Mikhail ESKAKOV and Alexey ISKAKOV. Equilibrium in secure strategies.
2012/62 Francis BLOCH and Axel GAUTIER. Strategic bypass deterrence.
2012/63 Olivier DURAND-LASSERVE, Axel PIERRU and Yves SMEERS. Sensitivity of policy simulation to benchmark scenarios in CGE models: illustration with carbon leakage.
2013/1 Pierre PESTIEAU and Maria RACIONERO. Harsh occupations, health status and social security.
2013/2 Thierry BRECHET and Henry TULKENS. Climate policies: a burden or a gain?
2013/3 Per J. AGRELL, Mehdi FARSI, Massimo FILIPPINI and Martin KOLLER. Unobserved heterogeneous effects in the cost efficiency analysis of electricity distribution systems.
2013/4 Adel HATAMI-MARBINI, Per J. AGRELL and Nazila AGHAYI. Imprecise data envelopment analysis for the two-stage process.
2013/5 Farhad HOSSEINZADEH LOTFI, Adel HATAMI-MARBINI, Per J. AGRELL, Kobra GHOLAMI and Zahra GHELEJ BEIGI. Centralized resource reduction and target setting under DEA control.
2013/6 Per J. AGRELL and Peter BOGETOFT. A three-stage supply chain investment model under asymmetric information.
2013/7 Per J. AGRELL and Pooria NIKNAZAR. Robustness, outliers and Mavericks in network regulation.

Books

G. DURANTON, Ph. MARTIN, Th. MAYER and F. MAYNERIS (2010), *The economics of clusters – Lessons from the French experience*. Oxford University Press.
J. HINDRIKS and I. VAN DE CLOOT (2011), *Notre pension en heritage*. Itinera Institute.
M. FLEURBAEY and F. MANIQUET (2011), *A theory of fairness and social welfare*. Cambridge University Press.
V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
J. HINDRIKS (2012), *Gestion publique*. De Boeck.

CORE Lecture Series

R. AMIR (2002), Supermodularity and complementarity in economics.
R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
A. SHAPIRO (2010), Stochastic programming: modeling and theory.